

Effective strategies for managing your research data - beginners session

Clair Castle, cmc32@cam.ac.uk
Lauren Cadwallader, lc340@cam.ac.uk

Office of Scholarly Communication
Research Data Management Facility

3 September 2018



Some housekeeping information:

1. Today: mixture of activities and us talking
2. Ask questions whenever you want
3. There is no fire alarm testing
4. Rule: everything you say remains confidential
5. Slides will be made available
6. Feedback form

Aims of today's workshop:

- Develop effective research data management skills
- Increase awareness of challenges around data management
- Improve awareness of significance of data management in context of reproducibility and research sharing
- Raise awareness of University data support services for researchers

Plan for today:

1: Data backup and file sharing strategies

2: File sharing basics

3: Data organisation

4: Managing personal and sensitive data

5: Data sharing whys and hows

To start with...

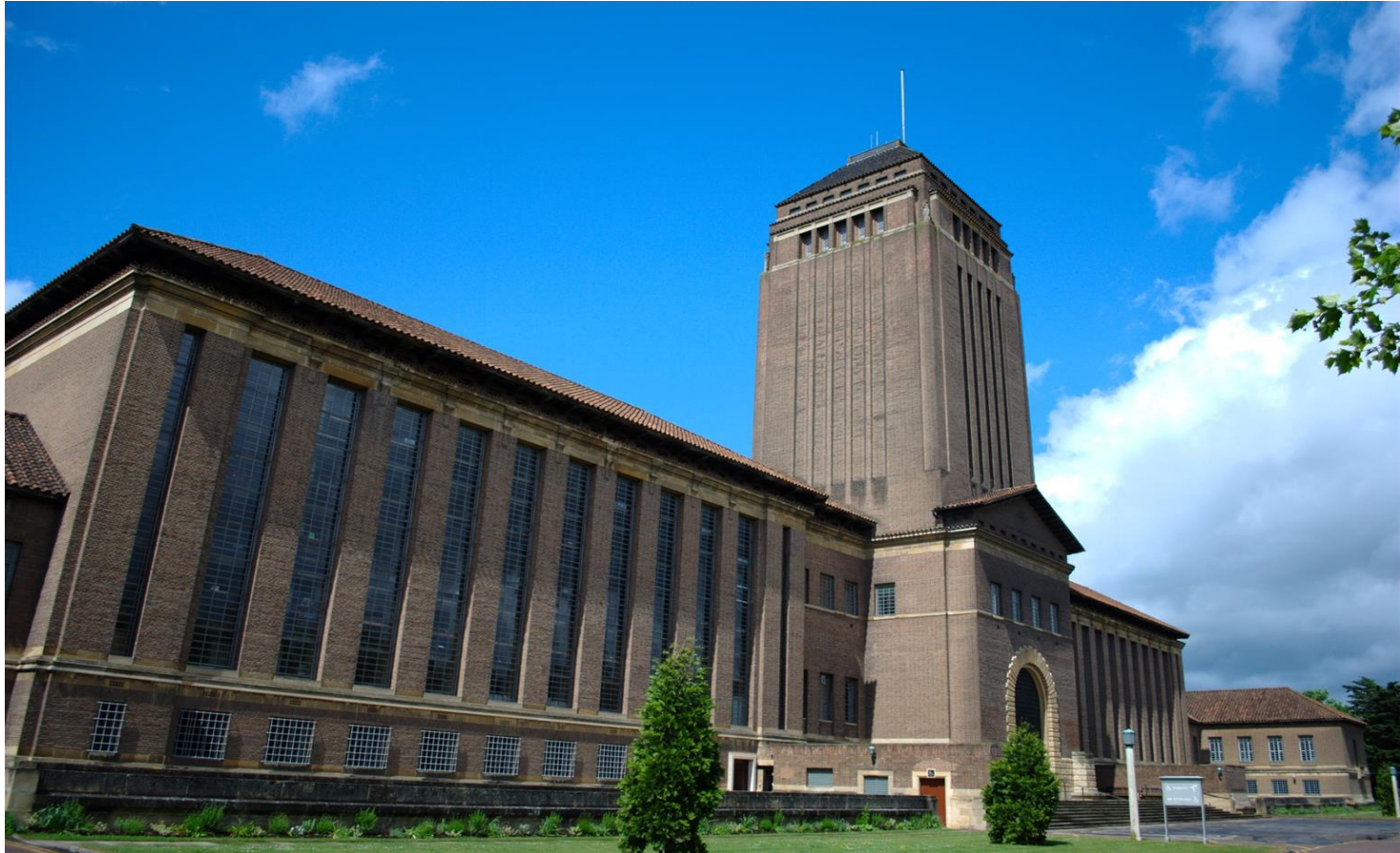
...let's get to know each other – talk to a person you don't know for 2 mins and exchange information:

- Who are you?
- Which department are you from?
- What type of research do you do?
- What kind of data do you collect?
- Why did you come here today?

After 2 minutes introduce each other to the group

Who are we?

The Office of Scholarly Communication



Copyright: <http://www.archdaily.com/408244/cambridge-university-library-landscape-design-competition>

Team of 16, headed by Dr Danny Kingsley, based at the University Library, a centre of digital preservation.

What does the Office of Scholarly Communication do?

We help researchers communicate their scholarly outputs



- Open Access
- Research Data Management
- Apollo
- Thesis Management
- Training

Scholarly Communication

Home Scholarly Communication Open Research Open Access Repository Copyright Monographs Theses Training Events

Find us on

Newsletters Twitter YouTube Blog

Tweets by @CamOpenAccess

Cambridge OA @CamOpenAccess
To Cambridge researchers: Do you know your Elements? [research-information.admin.cam.ac.uk/what-informati...](#)
Oct 5, 2017

Cambridge OA @CamOpenAccess
Embed View on Twitter

Events

12 OCT Python Training Sessions

17 OCT An introduction to Open Research (for PhD students in Humanities, Arts and Social Sciences)
[View all events >](#)

Advocating for the open sharing of research and data for the advancement of global knowledge

1 of 4

Open Access Research Data Theses Management

Open Access Research Data Theses Management

Monographs Author Tools Training



**What is research data
management?**

Research Data Management (RDM)



ORGANISING



STORING



ARCHIVING



SHARING

'Data' can mean a lot...

- Raw instrument readings
- Processed data
- Analysed data
- Genomic data
- Microscopic photos, western blot images and measurement
- Spreadsheets
- Videos
- Surveys and interviews
- Field notes
- Maps
- Lab books
- Physical samples
- Protocols

It's basically anything you produce in the course of your research!

FAIR data

Findable

- easily discoverable

Accessible

- no restriction to access

Interoperable

- doesn't rely on one specific type of software

Reusable

- is clearly licensed



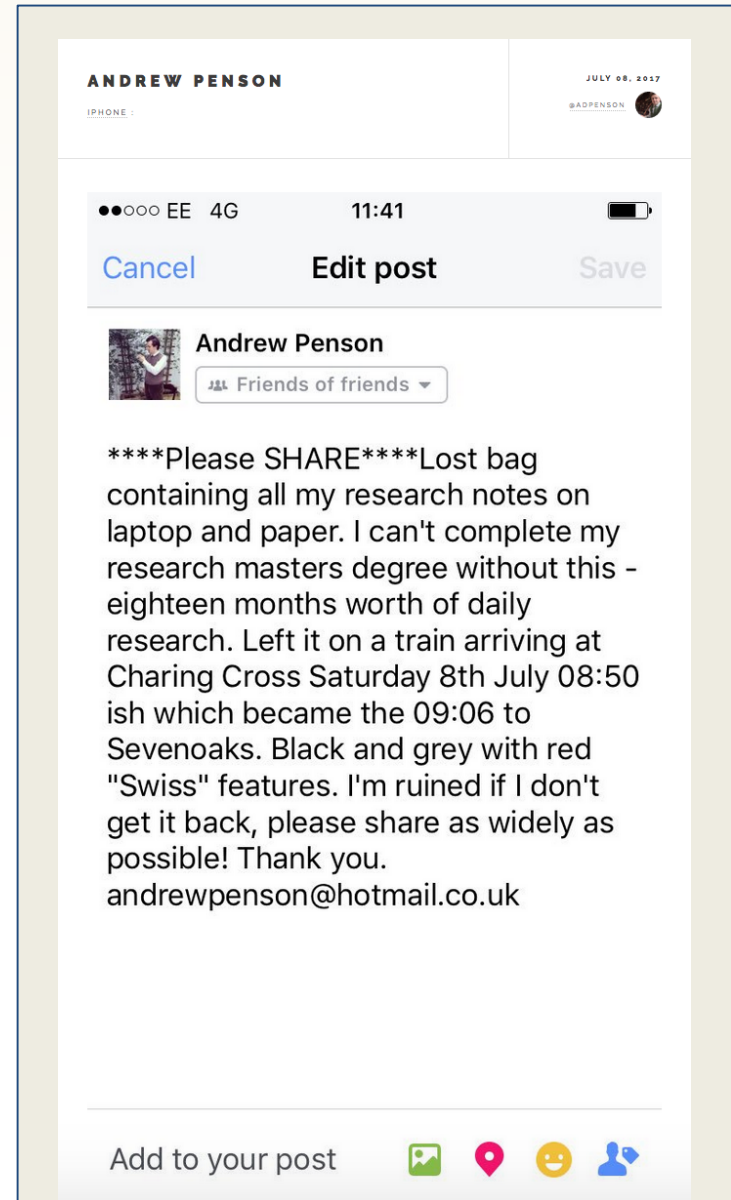
A photograph of a row of metal mailboxes mounted on a stone wall. The mailboxes are arranged in three rows. The top row contains mailboxes numbered 29 through 35. The middle row contains mailboxes numbered 36 through 42. The bottom row contains mailboxes numbered 43 through 49. A semi-transparent white rectangular box is overlaid on the middle row of mailboxes, containing the text '1. Data backup and file sharing strategies'. In the foreground, a portion of a concrete bench is visible.

1. Data backup and file sharing strategies

Disastrous data loss...



Credit: Peter Murray-Rust,
<http://blogs.ch.cam.ac.uk/pmr/2011/08/01/why-you-need-a-data-management-plan/>, August 2011, CC-BY





How much of your data would you lose if...?

How much of your data would you lose if...?

- your laptop got stolen
- your lab burnt down
- you lost your USB stick
- your portable hard drive got damaged
- Your stuff on Dropbox / Google Drive disappeared

How much data would you lose?

How could this data loss have been prevented?

Backup strategies:

- Departmental backup system
- External drives
- Online backups
- At least two backups, at two different locations

At least 2 backups at 2 locations:



Store at home!



Copy ASAP!



***Your
departmental
server***

Free software to manage backups (there is plenty of free software):

<http://www.2brightsparks.com/download-syncbackfree.html>



**Research
Data**

What is your name?

Hello, my name is...



Hello, my name is...
"FLASHDRIVE"



Hello, my name is "Hooray
for Mammoths"



Hello, my name is...
"LCadwallader(lc340)"



That's my name too!





2. File sharing basics

File sharing

- E-mail
- University of Cambridge solutions
<https://www.uis.cam.ac.uk/projects/storage-strategy/storage-services>
 - Research data storage, e.g. cold storage
 - Cloud storage, e.g. OneDrive
- (Secure) File Transfer Protocol – (S)FTP: set up by your IT support
- Local file servers
- Private cloud solutions – be cautious!




Beware of the cloud

Google services Terms of Use:

When you upload, submit, store, send or receive content to or through our Services, you give Google (and those we work with) a worldwide license to use, host, store, reproduce, modify, create derivative works (such as those resulting from translations, adaptations or other changes we make so that your content works better with our Services), communicate, publish, publicly perform, publicly display and distribute such content. The rights you grant in this license are for the limited purpose of operating, promoting, and improving our Services, and to develop new ones. This license continues even if you stop using our Services (for example, for a business listing you have added to

<https://www.google.com/intl/en/policies/terms/>

Cloud services comparison

			
Space included	1 TB	Unlimited	Unlimited
Price (per annum)	Free	£75 (+VAT)	Free
File history	90 days	30 days	Unlimited
Where are files stored?	Within UK	Within EU	Anywhere

Full comparison: <https://help.uis.cam.ac.uk/service/supporting-research/servers-data-storage-and-backup/data-storage/individual-storage>



3. Data organisation

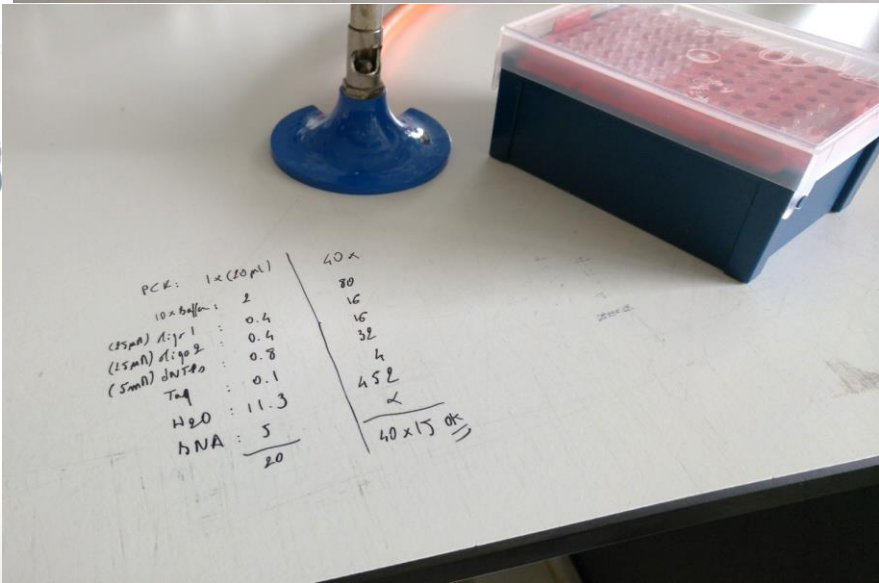
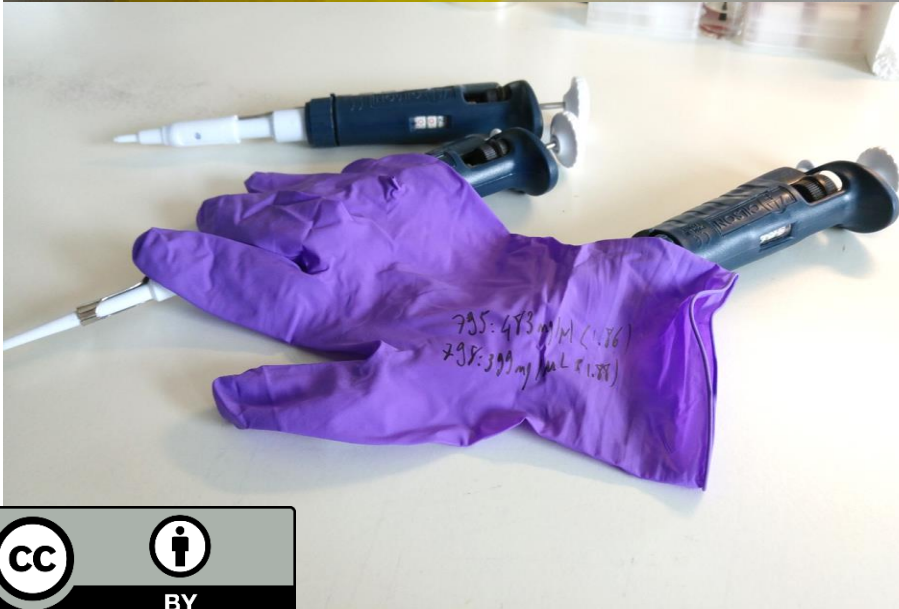
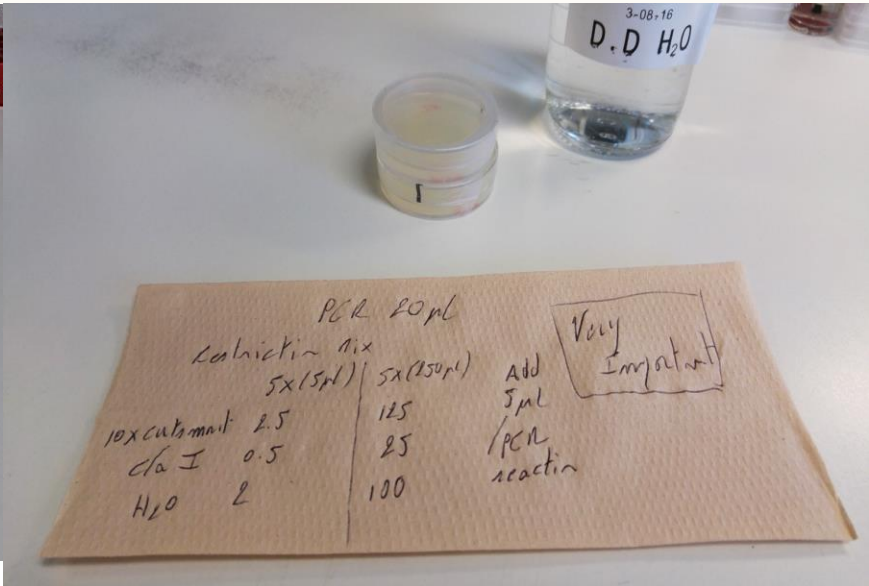
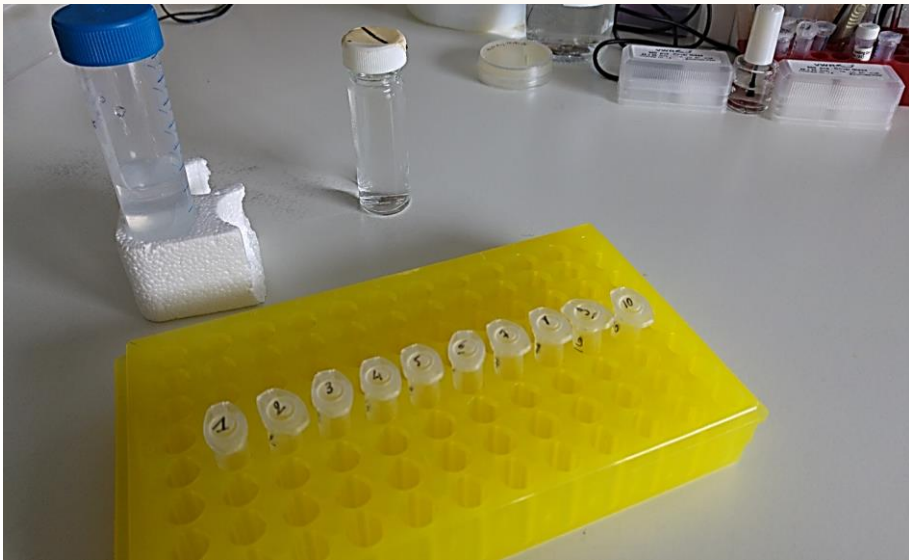
Physical samples



Lauren Cadwallader

Whether in the lab or field you need to think about good data management from the very beginning

Physical samples



Digitise your handwritten notes!

	A	B	C	D	E
1	Box #1				
2	Name of the sample	date in lab book	Samples left (if multiple samples)	Sample size	Notes
3					
4	Fibroblasts protein extract	23/03/2014	1	~40ul	Concentration: 0.6mg/ml
5	Fibroblasts protein extract	05/04/2014	7	~40ul	Concentration: 0.65mg/ml
6	Fibroblast protein extract for mass spec	19/08/2014	4	~50ul	Difficulties measuring concentration
7	Fibroblast RNA extract	20/09/2014	12	~10ul	Concentration: 80ng/ul
					Concentration: 100ng/ul; excellent prep
					Abcam: ab8227
					LOT: GR47300
					Concentration: 0.6mg/ml
					Rabbit polyclonal IgG

Samaca 2010 Registro de fotografías Página: ①

CEMENTERIO: 1002

Foto No.	Descripción	Fecha	Ini.
1	Foto board for panoramic shot SE → NE	26/7	CC
2	SE		
3	Panoramic shots.	GPS location	
4	over cemetery	of me taking	
5	to valley	shots:	
6	basin.	0432778	
7		8379375	
8	N	224m	
9	shot of huagüedo pit. GPS 0432778/8379383. View to N.		
10	Scattered bones on surface. GPS 0432784/8379370 227m		
11	Huagüedo pit with huagües. 0432823/8379370 224m		
12	Femur - sun damaged w/ fresh bit. 0432836/8379383 227m.		
13	View over E end of Cem to SSE, showing end of bluff 0432840/8379378 229m.		
14	Ceramic scatter. 0432832/8379374 229m		
15	Ceramic frags, one w/ handle. 0432791/8379392 230m		
16	Wooden posts on surface. 0432796/8379389 231m		
17	Snail scatter. 0432714/8379392 232m		
18	Juvenile remains & ceramic from cut C1.	✓	✓
19	Juvenile femur. Individual 3, sample 6, from cut C1.		
20	Comparison of femur (Ind 3) with white & folkers bgr. No board.		
21	Teeth in maxilla of Ind. 4. from cut C1		
22	Roots of deciduous molars Ind. 4		
23	Erupting teeth in maxilla. Ind. 4.		
24	Parietal frags of indiv. 5 from cut C1		
25	Parietal frag sample 9, indiv 5 ectocranial surface	✓	✓
26	Parietal frag sample 9, indiv 5 endocranial surface	✓	✓
27	Decorated ceramic from cut C1. Balsa 12.		

Lauren Cadwallader, PhD fieldwork
photo register

Physical samples need to be managed as well

- Create maps of your samples
 - can be simple Excel spreadsheets
 - and keep them up to date!
- Reference your samples:
 - dates in notebooks
 - supplier's name/code
- Add any relevant notes
- Think about converting paper copies to digital

Data organisation

Which example is better, and why?

Example A

Documents library

PhD data

- 25July
- Documents
- Experiments
- Experiments2
- From desktop
- Important
- Other
- PhD
- Talks
- Experiment 1
- Lab meeting FINAL
- Meeting notes
- My talk
- Paper submission
- PhD revised

Example B

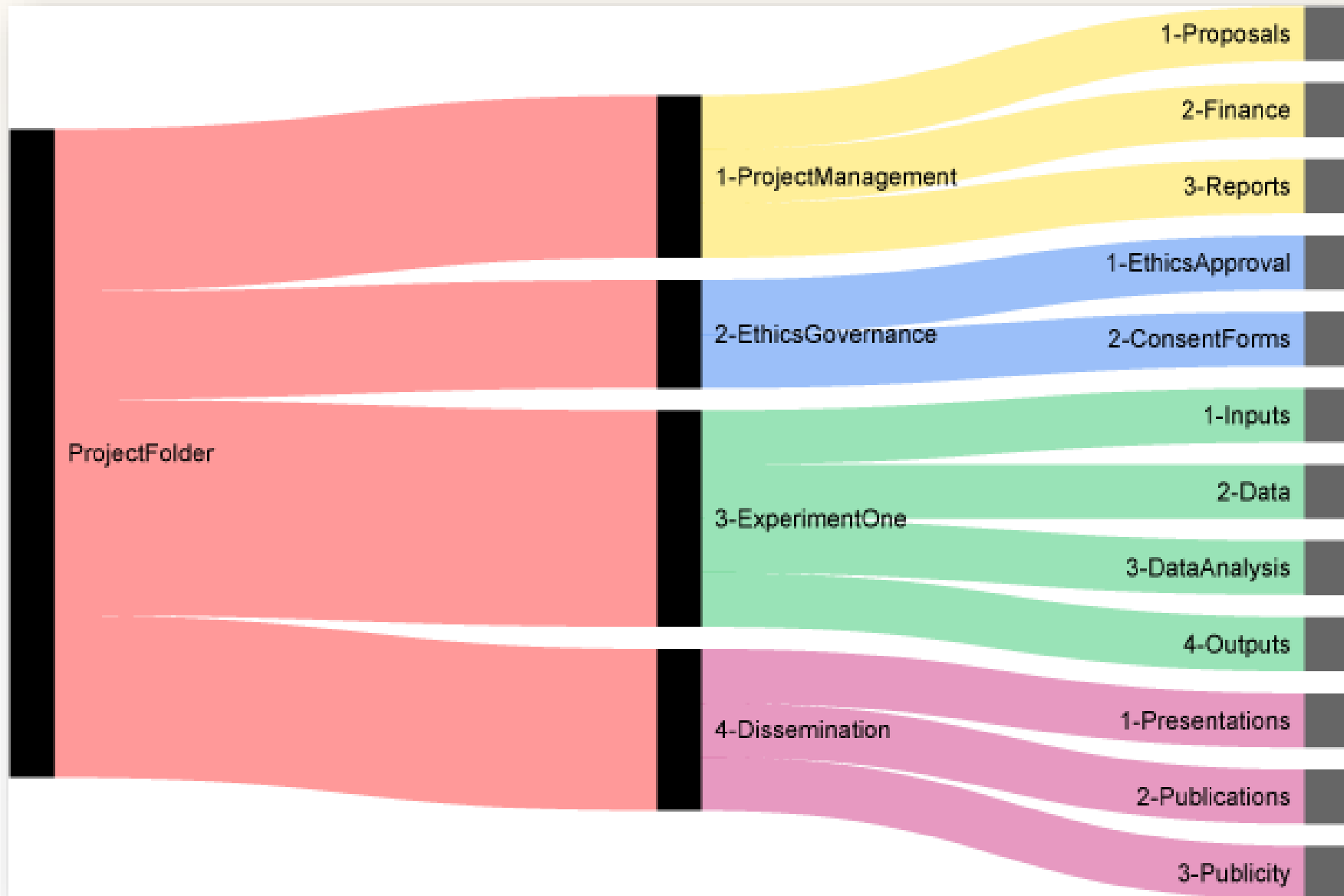
Documents library

PhD data2

- Conferences
- Downloaded publications
- Experimental data
- Financial documents
- PhD thesis
- Presentations
- Protocols
- Reagents
- Reports
- Training

Data organisation:

- Allow you to find files easily
- Meaningful to you and your colleagues
- Consistent



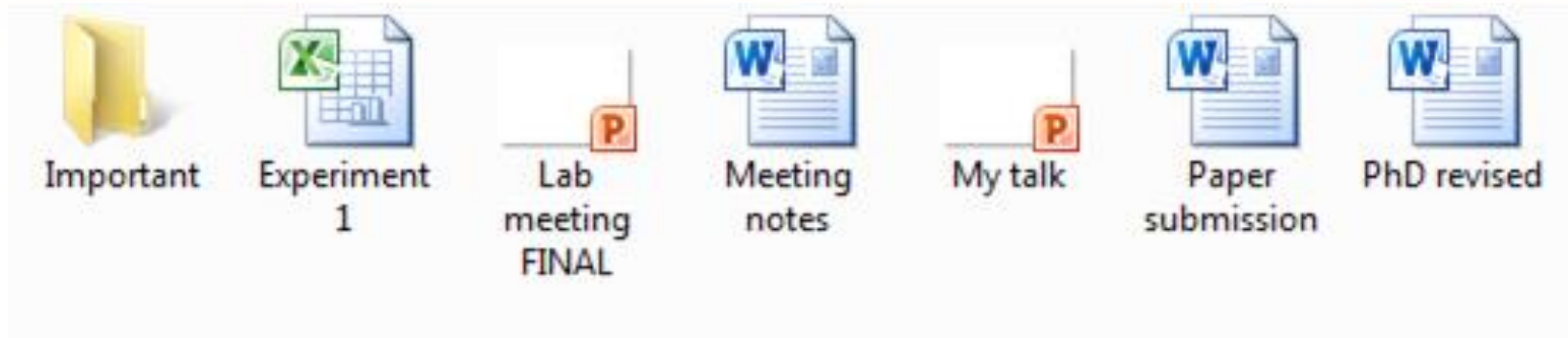
Copyright: http://nikola.me/folder_structure.html

File naming conventions – do they matter?



Copyright: <http://10pm.com/>

File naming conventions – do they matter?

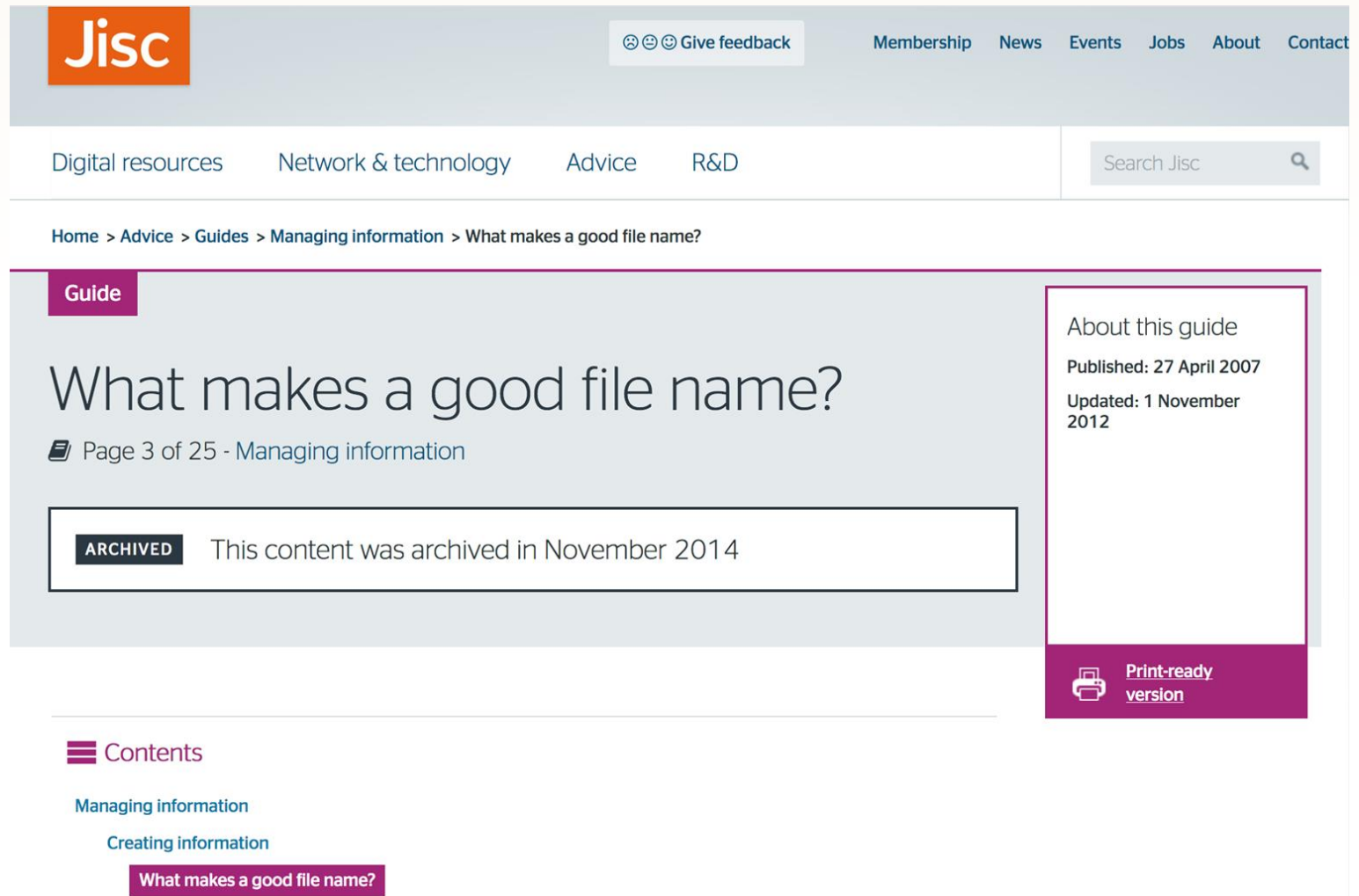


In 3 years time would you know what these are?



What do you think makes a
good file naming convention?

File naming conventions



The screenshot shows the Jisc website header with the Jisc logo, a 'Give feedback' button, and navigation links for Membership, News, Events, Jobs, About, and Contact. Below the header is a secondary navigation bar with links for Digital resources, Network & technology, Advice, and R&D, along with a search bar labeled 'Search Jisc'. The main content area displays a breadcrumb trail: Home > Advice > Guides > Managing information > What makes a good file name?. The guide title 'What makes a good file name?' is prominently displayed, followed by a sub-header 'Page 3 of 25 - Managing information'. A notice box states 'ARCHIVED This content was archived in November 2014'. To the right, a sidebar titled 'About this guide' provides publication details: 'Published: 27 April 2007' and 'Updated: 1 November 2012'. At the bottom right of the sidebar is a 'Print-ready version' button with a printer icon. A 'Contents' section on the left lists 'Managing information' and 'Creating information', with 'What makes a good file name?' highlighted as the current page.

Jisc

Give feedback

Membership News Events Jobs About Contact

Digital resources Network & technology Advice R&D

Search Jisc

Home > Advice > Guides > Managing information > What makes a good file name?

Guide

What makes a good file name?

Page 3 of 25 - Managing information

ARCHIVED This content was archived in November 2014

About this guide

Published: 27 April 2007

Updated: 1 November 2012

Print-ready version

Contents

Managing information

Creating information

What makes a good file name?

<https://www.jisc.ac.uk/guides/managing-information/good-file-name>

File naming conventions

A good file name enables its creator and anyone else within the institution to **identify** its content and context and to make a **decision about its relevance** without having to open the file itself.

In order to achieve this a file name should be:

- **Objective** e.g. “Project members Jones Lab” / “Coding project collaborators 2018”
- **Meaningful** e.g. “Lib Cttee” / “Library Committee”
- **Concise** e.g. “Conference report and other thoughts from Vienna” / “2018 Vienna Conference Report”
- **Standardised** e.g. follow agreed standards for personal names, dates, types of information, external bodies etc.

File naming conventions

TILS Document Naming Convention

Document naming for the TILS Division should follow this convention:

GDL_TILSDocNaming_V1_20090612.docx

A prefix shows the document type

The document title describes the content

The version number

The date in the format yyyyymmdd

http://www.data.cam.ac.uk/files/gdl_tilsdocnaming_v1_20090612.pdf

README files

If your description is long you can write a short description for the data record and then a full description as a **README file**.

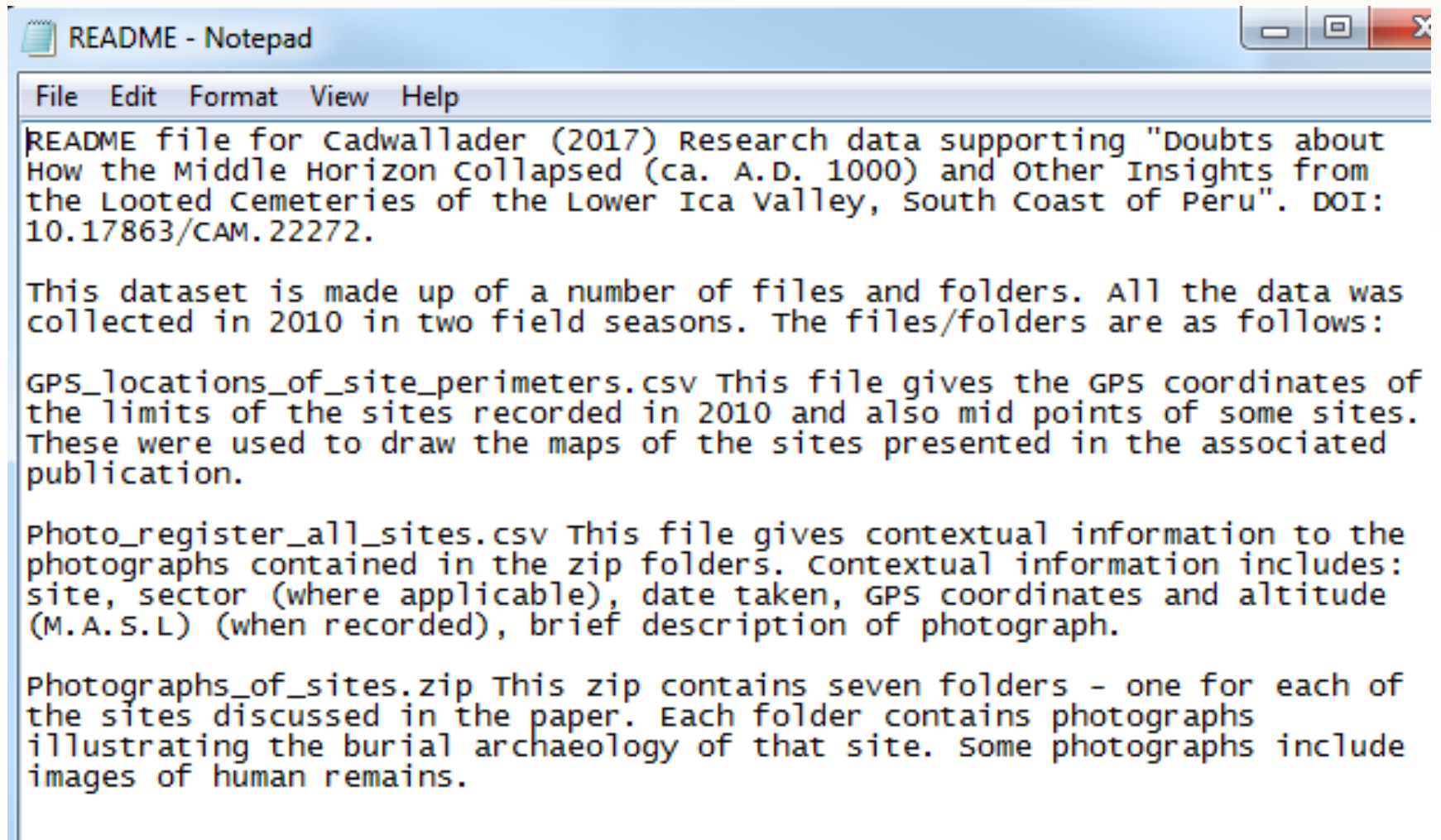
README files should contain:

- **General information** - title, authors, date of collection
- **File overview** - short description of the data each file contains and date it was created
- **Sharing information** - licence or restrictions placed on the data
- **Methodological information** - description of methods for data collection or generation, description of methods used for data processing
- **Data specific information** - variable list (including definitions) for tabular data, units of measurement, definitions for codes or symbols used to record missing data

Cornell University has excellent [README file guidelines](#).

Write as a plain text file.

README files



README file for Cadwallader (2017) Research data supporting "Doubts about How the Middle Horizon Collapsed (ca. A.D. 1000) and other Insights from the Looted Cemeteries of the Lower Ica Valley, South Coast of Peru". DOI: 10.17863/CAM.22272.

This dataset is made up of a number of files and folders. All the data was collected in 2010 in two field seasons. The files/folders are as follows:

GPS_locations_of_site_perimeters.csv This file gives the GPS coordinates of the limits of the sites recorded in 2010 and also mid points of some sites. These were used to draw the maps of the sites presented in the associated publication.

Photo_register_all_sites.csv This file gives contextual information to the photographs contained in the zip folders. Contextual information includes: site, sector (where applicable), date taken, GPS coordinates and altitude (M.A.S.L) (when recorded), brief description of photograph.

Photographs_of_sites.zip This zip contains seven folders - one for each of the sites discussed in the paper. Each folder contains photographs illustrating the burial archaeology of that site. Some photographs include images of human remains.

Writing metadata - examples

Description

This contains supporting **data needed to reproduce the DNS simulations reported in the accompanying paper**. The files include source code (in Fortran 90/MPI) and input files.

Description

Onchip Andreev devices and Ballistic Josephson junctions fabricated, and **measured by the authors at the Cavendish Laboratory, University of Cambridge UK, in the period Jan 2016 to April 2017. The measurements were done at low temperature (50 to 800 mK)**. The experimental methods are described in the associated publication.

Description

This data contains the corresponding MATLAB©-code for the numerical examples in the conference proceedings paper 'Gradient descent in a generalised Bregman distance framework'. **Download the zip-file and extract it to a folder of your choice. Execute the 'setpath.m' file to add all relevant files to the MATLAB© path**, and switch to the folder 'Examples'. This folder contains a script named 'phasereconstruction.m' that will compute the numerical examples as presented in the paper. **A detailed explanation of the script can be found in terms of the HTML-file 'phasereconstruction.html' in the sub-folder 'Manual'.**

Software

All calculations performed using the **development version of HANDE QMC** (homepage: <http://www.hande.org.uk/>). This will be reproducible using the publically available version of the HANDE QMC code after the next update (public github version: <https://github.com/hande-qmc/hande>)



4. Managing personal and sensitive data

What are personal data?

“records or other information that, on their own, or linked with other data or information in the possession of the data controller, can reveal the identity of an actual living person”

Data Protection Act (1998)

‘personal data’ means any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;

Art. 4 EU GDPR Definitions
(new law from 25th May 2018)

How to handle personal data?

- Don't collect it!
- Gain informed, preferably open and written, consent. See [UK Data Service](#) for examples.
- Anonymise data
 - Remove identifiers
 - Aggregate results
 - Generalise a variable
 - Remove outliers
- Use managed access repositories
- There are exemptions for research but you must add more detailed information to participant information sheets, and must not derive individual outcomes (e.g. drug treatment for specific person) but look at whole group

Guidance

University Research Ethics website:

<https://www.research-integrity.admin.cam.ac.uk/research-ethics>

University GDPR website:

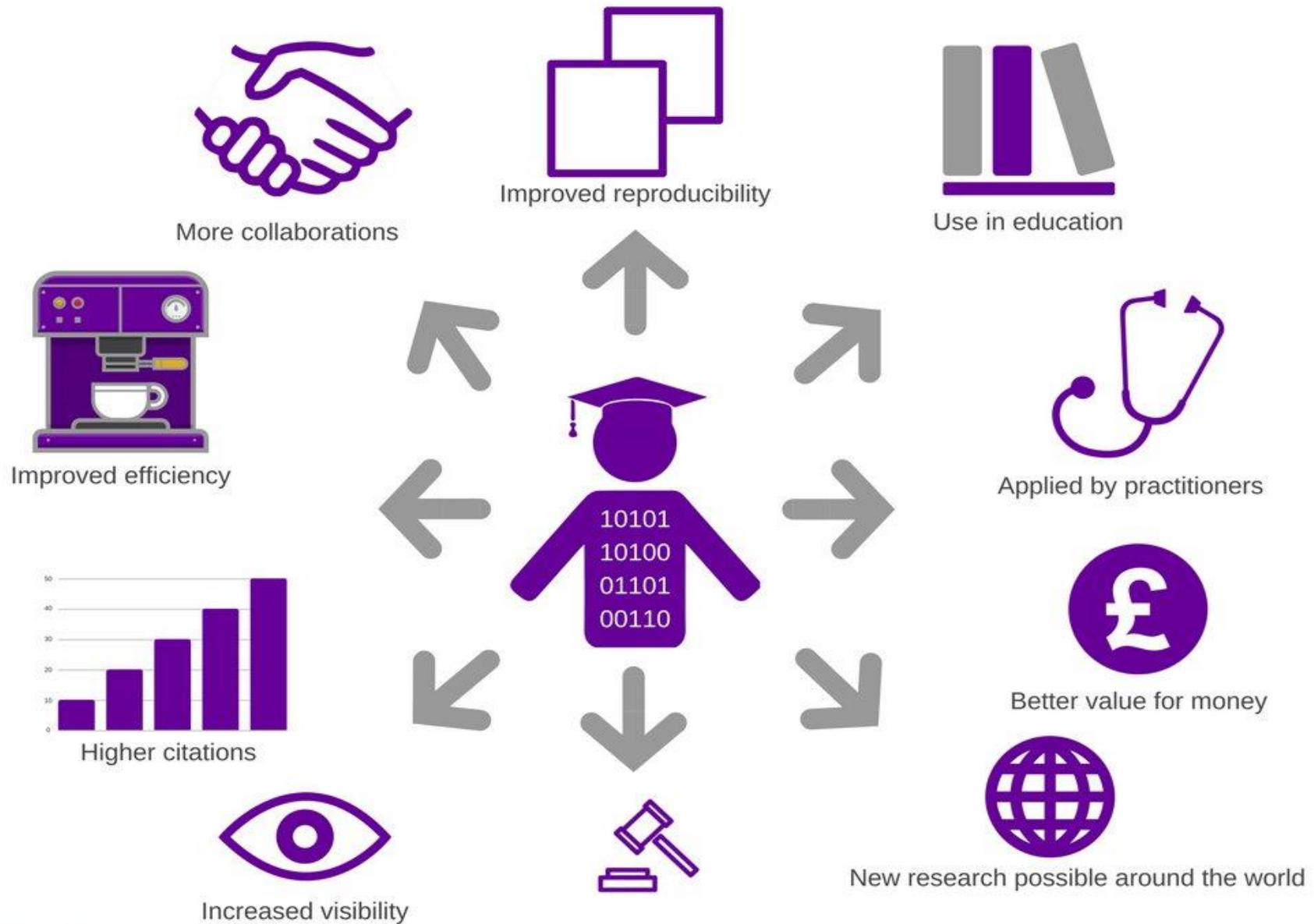
<https://www.information-compliance.admin.cam.ac.uk/data-protection/general-data-protection-regulation>

This is covered in more detail in our advanced course – find out more later.

A close-up photograph of a person's hands holding a rustic, reddish-brown ceramic bowl filled with cherry tomatoes. The tomatoes are mostly red, with one green one on the left. The person's left hand is wearing a light-colored, textured gardening glove with blue polka dots. The background is a blurred green field with a yellow flower. A semi-transparent white rectangular box is overlaid on the center of the image, containing the text '5. Data sharing whys and hows'.

5. Data sharing whys and hows

Sharing your research is A Good Thing



Open Access to research data

“Publicly funded research data are a public good (...), which should be made openly available with as few restrictions as possible...”



<https://www.ukri.org/funding/information-for-award-holders/data-policy/common-principles-on-data-policy/>





OPEN

Write down 2-3 things that worry you about sharing your data openly

Open Access to research data

What if:

- Someone steals my idea
- It stops me from being published
- It's too much work
- It's my data and I don't want to share it
- I have to pay for it



Katie Hughes



Funder names arranged alphabetically. Click on the hyperlink to see the full-length policy.

BBSRC Read the policy	<p>Adherence to data management plan will be monitored and built into the Final Report score, which may be taken into account for future proposals.</p> <p>Research data that supports publications must be stored for 10 years.</p> <p>Grantholders are requested to capture and record data sharing activities, including details of where and how data have been shared, in the appropriate places on ResearchFish.</p> <p>Detailed guidelines about BBSRC requirements are available here.</p> <p>We have also discussed BBSRC policy directly with Michael Ball from the BBSRC. Our discussion and resulting clarifications of the BBSRC policy are published here.</p>	June 2018
British Heart Foundation Read the policy statement	<p>"Safeguards should be in place to respect the confidentiality of patients, while also ensuring that medical researchers can gain access to patient data within a secure environment."</p>	June 2018
Cancer Research UK Read the policy	<p>Any applicants who consider that the data arising from their proposal will not be suitable for sharing must provide clear reasons for not making it available.</p> <p>Investigators carrying out research involving human participants must ensure that consent for data sharing is obtained from participants; research data should be anonymised prior to sharing.</p> <p>Research data should be available for sharing for a minimum period of five years from the end of a research grant.</p> <p>CRUK also issued a list of FAQ on data sharing.</p> <p>We have also invited JCRUK to discuss their data sharing requirements with researchers at Cambridge. We have published blog posts</p>	June 2018

Key policy highlights

Date the policy was last checked or updated.

Cambridge data repository




www.data.cam.ac.uk

We will check the data, upload it into the repository and send a link to it



And each submission gets its own record

Research data supporting "Papers, policy documents and patterns of attention"



Citation
Cadwallader, L., & Altmetric.com. Research data supporting "Papers, policy documents and patterns of attention" [Dataset]. <https://doi.org/10.17863/CAM.4584>

Description
This is the Altmetric.com data for the set of journal articles used in this research. The data was provided by Altmetric.com, a research metrics company who track and collect the online conversations around millions of scholarly outputs. Altmetric continually monitors a variety of non-traditional sources to provide real-time updates on new mentions and shares of individual research outputs, which are collated and presented to users via Altmetric.com. The data was collated on the 15/08/2016. Any subsequent adjustments to the original data have been made by Dr Lauren Cadwallader and are fully explained in the document.

Software
Any Excel compatible software


Keywords
altmetric, policy, impact, online attention, journal article

Relationships
Publication Reference:
<https://doi.org/10.17863/CAM.4844>

Sponsorship
Altmetric.com

Identifiers
This record's DOI:
<https://doi.org/10.17863/CAM.4584>

View / Open Files
[README.pdf \(PDF, 82Kb\)](#)
[RAW_AltmetricCadwallader_20160815.csv \(Unknown, 18Kb\)](#)
[DATA_AltmetricCadwallader_NoForm ulae_20160926.xlsx \(Microsoft Excel 2007, 152Kb\)](#)
[DATA_AltmetricCadwallader_20160926.xlsx \(Microsoft Excel 2007, 196Kb\)](#)




Authors
 Cadwallader, Lauren
Altmetric.com

Publication Date
2016-09-23

Publisher
University of Cambridge

Type
Dataset

Metadata
[Show full item record](#)

  Blogged by 2
 Tweeted by 7

[See more details](#)

Rights
Attribution 4.0 International
Licence URL:
<http://creativecommons.org/licenses/by/4.0/>

Citation

Reference to
other outputs

Licence

DOI!

<https://doi.org/10.17863/CAM.4584>



Google indexes the repository!



Research Data Supporting "Contagious risk taking: social information and c



All

News

Images

Videos

Shopping

More ▼

Search tools

About 180 results (0.78 seconds)

Contagious risk taking: social information and context ... - Nature

www.nature.com › [Scientific Reports](#) › [Articles](#)

by AL Greggor - 2016 - [Cited by 1](#) - [Related articles](#)

10 Jun 2016 - All **data** were analysed in R. Birds were deemed to have access to a S. R., Rutter, J. E. & Tonra, C. M. A call for full annual cycle **research** in A.T. was **supported** by a BBSRC David Phillips Fellowship (BB/H021817/1).

Research Data Supporting "Contagious risk taking: social information ...

<https://www.repository.cam.ac.uk/handle/1810/256096> ▼

by AL Greggor - 2016

26 May 2016 - **Research Data Supporting "Contagious risk taking: social information and context influence wild jackdaws' responses to novelty and risk"**.

[PDF] Contagious risk taking - University of Cambridge

https://www.repository.cam.ac.uk/.../Greggor_et_al-2016-Scientific_Reports-VoR.pdf... ▼

by AL Greggor - 2016 - [Cited by 1](#) - [Related articles](#)

10 Jun 2016 - ... of novel food avoidance provide empirical **support** for the suggestion that ... for populations of other passerine species when the **data** is compared Guidelines for the Treatment of Animals in Behavioural **Research** and.

Average numbers of visits

84,000 downloads per month!!!



~34,000 visitors per month

Cambridge support for data management and sharing



Research data services provided by the Cambridge data team:

- Data repository
- Data Management Plan checking
- Advice on funders' policies
- Training on research data management
- Online information
- On demand data management consultancy

www.data.cam.ac.uk

Research Data Management

Home

Data Management Guide

Support

Data Repository

Research Data Policies

FAQ

News

Ev



UPLOAD YOUR DATA

Have a question? E-mail info@data.cam.ac.uk

Training and support

03
SEP

Effective strategies for managing your research data (beginners session)

Monday, 3 September, 2018 at 10:00-12:00

8 Mill Lane, Lecture Room 5

05
SEP

Effective strategies for managing your research data (advanced session)

Wednesday, 5 September, 2018 at 14:00-16:00

8 Mill Lane, Lecture Room 10

11
SEP

Reproducing and reusing research code: lunch and learn with Code Ocean

Tuesday, 11 September, 2018 at 12:30-13:30

Postdoc Centre, Newman Library @ Biomedical Campus



Researchers and students can now not only make their code and data available for their academic papers, but also enable others to reproduce the results with a single-click.

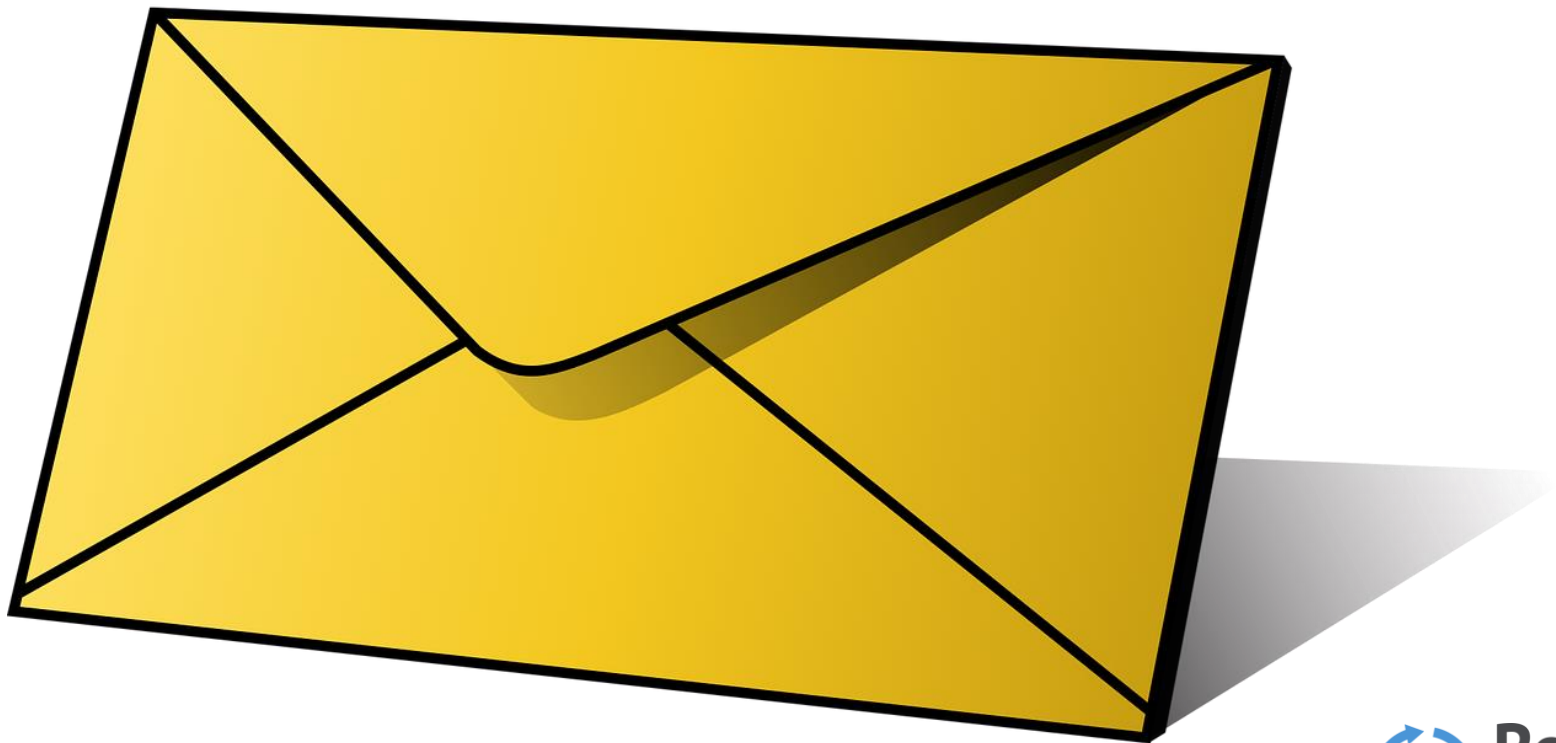
Code Ocean is an easy-to-use executable repository and reproducibility platform that facilitates replication and reuse of research code. This demo will provide an overview of the Code Ocean platform and explore benefits such as:

- **preservation** - code will work today, tomorrow, next week, next year
- **advanced tech** - suite of tools which follow reproducibility best practices
- **impact** - enable easy reuse of code to extend research
- **collaboration** - code is easy to share and discover.

Lunch will be provided during the session, which will include a demo and discussion.

You may find it useful to bring your own internet-enabled device to the session

Your data management resolution for
the next month



Your data management resolution for the next month

1. YOU (today)

- a) Address an envelope to yourself
- b) Write your data management resolution for the next month on a “post-it note” and place it in the envelope
- c) Write today’s date on the front of the envelope

2. US:

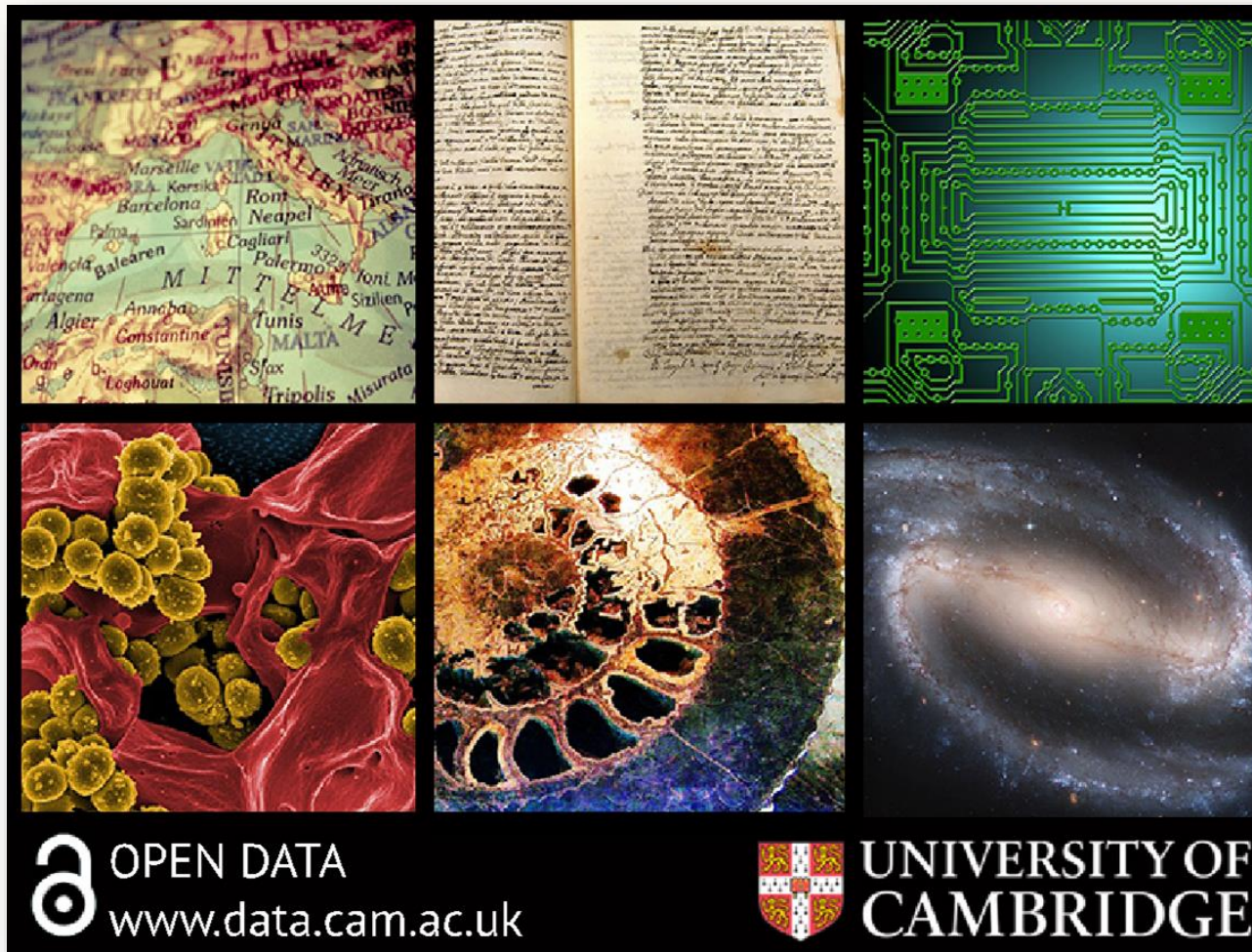
- a. We will send you the envelope on 3 October

Want to enhance your basic RDM skills?

Attend our advanced course!

<https://www.training.cam.ac.uk/osc/event/2623409>

Take-home message:



www.data.cam.ac.uk
info@data.cam.ac.uk

Thank you

Enquiries about research data: info@data.cam.ac.uk

www.data.cam.ac.uk



@CamOpenData